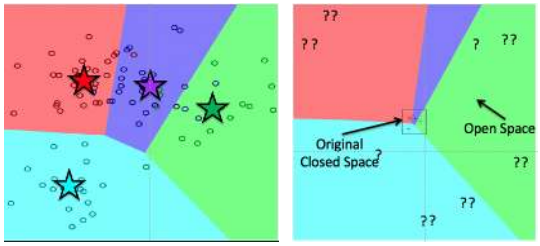
Learning and the Unknown: Surveying Steps toward Open World Recognition

T. E. Boult,[1] S. Cruz,[1] A.R. Dhamija,[1] M. Gunther,[1] J. Henrydoss,[1] W.J. Scheirer[2]
[1]University of Colorado Colorado Springs, Colorado Springs, CO 80918
[2]University of Notre Dame, Notre Dame, IN 46556

Abstract

As science attempts to close the gap between man and machine by building systems capable of learning, we must embrace the importance of the *unknown*. The ability to differentiate between known and unknown can be considered a critical element of any intelligent self-learning system. The ability to reject uncertain inputs has a very long history in machine learning, as does including a *background* or *garbage* class to account for inputs that are not of interest. This paper explains why neither of these is genuinely sufficient for handling unknown inputs – uncertain is not unknown, and unknowns need not appear to be uncertain to a learning system. The past decade has seen the formalization and development of many *open set* algorithms, which provably bound the risk from unknown classes. We summarize the state of the art, core ideas, and results and explain why, despite the efforts to date, the current techniques are genuinely insufficient for handling unknown inputs, especially for deep networks.



(a) Example four-class model from closed set point of view. (b) Zooming out to show some open space.

Figure 1: The issue of open space can be seen by zooming out from around the training data. Open space is the region far from training samples. A traditional classifier, e.g., NCM shown here, will label everything including the unknown "?" inputs. Even points infinitely far away are labeled.

1 Introduction

"Intelligence comes with hard work and curiosity for the unknown." - Roberto Llamas

With the advent of rich classification models and high computational power, recognition systems are finding many operational applications. Recognition in the real world poses multiple challenges that are not apparent in controlled lab environments. At prediction time, an operating system has to deal with myriad unseen categories. Consider the goal of "A.I." for autonomous driving. While we might train such a system with terabytes of data, it is impossible to anticipate and train with all possible inputs. However, with a critical safety system, even a small fraction of errors on unknown inputs could be, quite literally, deadly. Most real data is inherently dynamic, and the world unpredictable; novel inputs/categories must be handled by designing systems that can ignore/reject them or designing systems that continuously detect novel inputs and do something with unknown inputs. Ideally, the system needs to label and add novel detected objects as a new item to be learned.

Early A.I. systems and papers often involved the "closed world assumption," i.e. the system model was complete, and the system could reason using what was observed as well as what was not. Even back in the early 80s (Hewitt and Jong 1983) noted, "At first glance, it might seem that the closed world assumption, almost universal in the A.I. and database literature, is smart because it provides a ready default answer for any query. Unfortunately, the default answers provided become less realistic as the Open System increases in size." The closed world assumption led to fragile systems that failed, and so fell out of favor as researchers tried to move out of the lab and into the open world.

To help visualize the key issue of open set recognition (OSR), consider the four-class problem shown in Fig. 1a, with a Nearest Class Mean (NCM) model (Mensink et al. 2012), where the star is the NCM. Then consider what is labeled when zoomed out (Fig.1b).

By the turn of the century, most A.I. systems stopped explicitly exploiting the *closed world assumptions*, many moving to probabilistic Bayesian reasoning. More recently systems have turned to using learning-based models. Unfortunately, almost all machine-learning-based system have implicitly continued to make that assumption because they classify all inputs into one of their training classes. Similarly, Bayesian reasoning implicitly retains the closed world assumption. Research often seek closed set classifiers, that approximate the Bayesian optimal posterior probability,

$P(C_l|x'; \mathcal{C}_1, \ldots, \mathcal{C}_M), l \in \{1, \ldots, M\}$, for a fixed set of classes, where x' is an input sample, l is the index of class $\mathcal{C}_l$ (a particular known class), and M is the number of known classes. When Ω unknown classes appear at query time, however, the Bayesian optimal posterior becomes $P(C_l|x'; \mathcal{C}_1, \ldots, \mathcal{C}_M, U_{M+1}, \ldots, U_{M+\Omega})$, which cannot be computed/modeled because classes U_{M+1} through U_Ω are unknown. Even the core law of total probability, essential in Bayes theorem, cannot be applied unless one presumes the probability of all unknown inputs is known. In other words, using Bayes theorem requires implicitly making a closed world assumption. For classifiers that assess confidence in terms of signed distance from a decision boundary, or some calibration thereof, this misclassification will occur with high confidence if the unknown is far from any known data — a result that is very misleading.

Until recently, almost all evaluations of machine-learning-based recognition algorithms have implicitly been "closed set" whereby the system is only tested on classes known at training time. A more realistic scenario for applications is accepting that the world is an open set of objects, that our knowledge is always incomplete, and thus that unknown classes should be submitted to an algorithm during testing. This paper reviews and extends the formalizations of **open set recognition** (Scheirer et al. 2013; Scheirer, Jain, and Boult 2014) and **open world recognition** (Bendale and Boult 2015), collectively open recognition.

Related, but formally separate from open recognition, are the approaches that use classifiers with rejection (Chow 1970; Matan et al. 1990; Fumera and Roli 2002; Rong and Metaxas 2006; Bartlett and Wegkamp 2008; Grandvalet et al. 2008), novelty, anomaly or outlier detection (Hodge and Austin 2004; Markou and Singh 2003). These can help with rejecting unknown inputs but lack the formal properties of provably bounded open space risk. This paper briefly reviews these areas, recent results, and explains the difference between them and open recognition.

As highlighted in a recent *AI Magazine* piece (Dietterich 2017), OSR is a growing subfield that that is critical for robust systems. In the five years following our formalization of OSR (Scheirer et al. 2013), the problem has received significant attention. The work in this area has received hundreds of citations from a wide range of application areas that need, or are using, OSR including:

- Audio analysis (Battaglino, Lepauloux, and Evans 2016; Krstulović 2018),
- Automatic Target Recognition (Scherreik and Rigling 2016b; 2016a; Roos and Shaw 2017),
- Autonomous Navigation/Mobile Robitcs (Zamora and Yu 2016; Sünderhauf et al. 2016),
- Biomtrics (Chiachia et al. 2014; Pinto et al. 2015; Rattani, Scheirer, and Ross 2015; Juefei-Xu and Savvides 2016; Günther et al. 2017; Perera and Patel 2017; Bao et al. 2018),
- Cyber Intrusion/Malware Detection (Henrydoss et al. 2017; Cruz et al. 2017; Rudd et al. 2017),
- Data Fusion (Neira et al. 2018),
- Domain Adaption (Gopalan et al. 2015; Busto and Gall 2017),
- Forensics (Costa et al. 2014; Rocha et al. 2017; Navarro et al. 2018),
- Lifelong Learning (Chen and Liu 2016; Rebuffi et al. 2017),
- Natural Language (Prakhya, Venkataram, and Kalita 2017; Doan and Kalita 2017; Grave, Cisse, and Joulin 2017),
- Novelty Detection (Bodesheim et al. 2015; Lazzaretti et al. 2016; Schultheiss et al. 2017),
- Package Authentication (Schraml et al. 2017),
- Unmanned Aerial Systems and Aerial Imagery (Poitevin, Pelletier, and Lamontagne 2017; Bapst et al. 2017), and
- Zero-shot learning (Lampert, Nickisch, and Harmeling 2014; Chao et al. 2016; Xian, Schiele, and Akata 2017; Fu et al. 2018; Xian et al. 2018)

There have also been papers developing their own models/algorithms for OSR: (Cardoso, França, and Gama 2015; Liu et al. 2016a; Ferreira and Giraldi 2017; Zhang and Patel 2017; Mu, Ting, and Zhou 2017; Günther et al. 2017; Neal et al. 2018; Bansal and Weld 2018).

When using classic "explainable" features for a problem, the problem of open recognition can be well formulated in either image or feature space. With the shift to deep networks, which combines learning features and learning the classifier, the problem becomes more difficult and is still largely unsolved. The paper ends with a discussion of some of the issues that make open set deep learning so tricky, and so vital as we move toward intelligent systems.

At the core of intelligence is the ability to recognize when we do not know something, analyze the need to learn about it, and then, when needed, to adapt and learn it – a process we call open world learning. As research moves towards building intelligent systems and deploying machine-learning-based systems, this is a critical but understudied area of research. The fundamental research question addressed in this paper is how to formally address the unknown in machine learning systems. This paper reviews formal approaches for handling those issues, but the problem is far from solved. In our ever-changing world, we recommend you *do not trust a claim of intelligence that does not admit when it does not know or does not continue to learn.*

2 Formalizing Open Set Recognition

We introduced the first formalization of open recognition in (Scheirer et al. 2013) with essential properties: bounding the open space risk and ideally balancing it with empirical risk. Empirical risk, measured on training data, is easy to define and practical to optimize, but how to extend the model to capture the risk from unknown inputs is the critical issue for OSR. That paper argued the essential element of OSR is minimizing the volume of space representing the learned recognition function f, outside the reasonable support of the positive samples, because clusters of unknown samples in initially unlabeled regions are more likely to be negatives and increase the recognition error. Note, this is entirely different from the classic binary classifier approach, which tries to label everything, and often generalization to infinite space in some directions. Let's proceed to formalize this idea. Let f be a measurable recognition function over input space $\mathcal{X}$, for known valid class $\hat{V}$. Let $S_{\hat{V}}$ be a union of balls of radius r_o that includes all of the training examples for all known $x \in \hat{V}$, let $\mathcal{O}$ be the open space with $\mathcal{O} \subset= \mathcal{X} - \mathcal{S}_{\hat{V}}$. Open space risk $R_{\mathcal{O}}(f)$ for class $\hat{V}$ can be defined as $R_{\mathcal{O}}(f) = \frac{\int_{\mathcal{O}} f_{\hat{V}}(x)dx}{\int_{S_{\hat{V}}} f_{\hat{V}}(x)dx}$, where open space risk is considered to be the relative measure

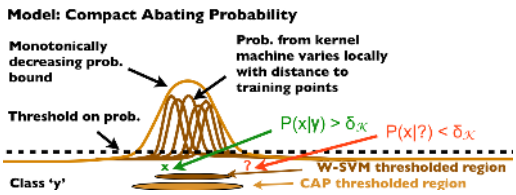


Figure 2: Compact Abating Probability (CAP) model which, when thresholded, can bound open space risk. CAP models are often used in novelty, anomaly, or outlier detectors.

of positively labeled open space compared to the overall measure of positively labeled space. Following (Scheirer et al. 2013) we can formally define the OSR problem as follows:

Definition 1. The Open Set Recognition (OSR) Problem: *Using training data with positive samples $v_i \in \hat{V}$, and other known class samples $k_j \in \hat{K}$, and an empirical data error/accuracy measure $\mathcal{E}$, find a measurable recognition function $f \in \mathcal{H}$, where $f(x) > 0$ implies positive recognition for class $\hat{V}$, and f is defined as minimizing the OSR error:*

$$\underset{f \in \mathcal{H}}{\operatorname{argmin}} \{R_{\mathcal{O}}(f) + \lambda_r \mathcal{E}(f(v_i); f(k_j))\} \tag{1}$$

subject to

$$m\alpha \leq \sum_{i=1}^{m} \phi(f(v_i)) \quad \text{and} \quad n\beta \geq \sum_{j=1}^{n} \phi(f(k_j)) \tag{2}$$

where λ_r specifies the regularization tradeoff between open space risk and empirical risk, where $\alpha \geq 0$ and $\beta \geq 0$ allow a prescribed limit on true positive and/or false positive rates, and $\phi(z)$ is a given loss function, e.g. the classic soft margins hinge loss $\phi(z) = max(0, 1 - z)$ or squared hinge loss $\phi(z) = max(0, 1 - z)^2$ functions.

The full regularized optimization for OSR error can be difficult. However, since empirical risk is always bounded, it is straightforward to see that if $f(x) > 0$ for a infinite amount of space, then $R_{\mathcal{O}}(f) = \infty$ and such an f cannot even be an approximately optimal solution because its open set risk is unbounded while the function $f(x) = 0$ always has bounded OSR error. For that reason we consider a minimal requirement for an algorithm to be called a formal OSR algorithm to be $R_{\mathcal{O}}(f) < B$ for a finite bound B. Any function that realizes such a finite OSR error will be within some constant of the optimal error. In our various papers on the topic, (Scheirer, Jain, and Boult 2014; Bendale and Boult 2015; Júnior et al. 2016; Rudd et al. 2018), we show multiple ways to find algorithms with bounded open space error.

3 Extensions: CAP and Open World Recognition

In (Scheirer, Jain, and Boult 2014), we introduced the idea of a *Compact Abating Probability* or CAP model, and showed that such models always have a bounded OSR error. The basic idea of a CAP model, see Fig. 2, is that if the region of support for the classifier is decreasing in all directions away from the training samples, then thresholding it will bound the open space risk. With the theorems in that paper, we showed that some models already in the literature, such as a thresholded RBF one-class Support Vector Machine (SVM), formally bound open-space risk. That paper also introduced the WSVM, which was far more accurate than any prior OSR algorithm. The WSVM uses Extreme Value Theory (EVT) to calibrate and combine multiple kernel-SVMs with a CAP model.

In an attempt to reduce the computational cost compared to the WSVM, we developed another variant, the Pi-SVM (Jain, Scheirer, and Boult 2014). While it was generally as accurate and sometimes more accurate than the WSVM, at a small fraction of the cost, we could not prove that the algorithm always had bounded open space risk. Examining why, we recognized that the same issue occurred for a regular SVM, and proved that an RBF-SVM with a CAP consistent kernel has bounded open space risk *if and only if*, all its bias terms are negative. The negative bias property might also be why some SVM-based systems do not have a problem with unknown inputs – if the data happens to result in kernels with negative bias terms, then the SVMs will have bounded open space risk for that data. We developed an SVM variant, the specialized SVM (Júnior et al. 2016), which ensures negative bias and hence OSR. While any decreasing function can be a CAP model, and many existing novelty, anomaly, or outlier detectors use them, the modeling of the tails of the distribution, the extreme values, is very critical to balancing open set risk and classification accuracy. That is why most of our work turned to EVT.

In the follow-on work (Bendale and Boult 2015) we formally defined and extended the OSR problem to *open world recognition*. An effective open world recognition system must efficiently perform four tasks: detect unknowns, choose which points to label for addition to the model, label the points, and update the model. While something like the WSVM could formally be used for incremental learning, since there are incremental SVM tools (Caragea, Silvescu, and Honavar 2000), those algorithms do not scale well. A simple and more efficient algorithm, nearest non-outlier (NNO), was proposed as part our open world recognition approach. The proofs of (Scheirer, Jain, and Boult 2014) were extended to handle transformations of CAP models, e.g., combinations of multiple cap models also bound open space risk, and thus NNO offered a formal solution to open world recognition. Those proofs also allowed us to show that thresholding many widely used density-based models such as Guassian Mixture models (GMMs) or Kernel-Density Estimator (KDE), are also provably OSR algorithms. Unfortunately, NNO was not very accurate, as the algorithm used thresholded distances from the nearest class mean and otherwise ignored distributional information. Weak classifiers are a persistent problem for incremental learning: it is not immediately obvious how one might extend class boundary models from classical machine learning theory, e.g., kernel machines, to incorporate both incremental learning and open set constraints.

In (Rudd et al. 2018), we developed the Extreme Value Machine (EVM), a new non-linear radial basis function approach that supports both high-quality non-linear decision

boundaries and efficient incremental learning. We proved that the EVM has bounded open space risk, and showed it was significantly more accurate than NNO on large-scale Image-Net testing using deep features.

4 Approaches that sometimes solve OSR

In this section, we discuss approaches some people argue address OSR and explain why these algorithms generally have an infinite open set risk.

A common question is why a pure detection system with a binary output, e.g., a face detector, is not a solution, and why it does not already limit its open set risk. It turns out that for some detection systems, in particular, those that use GMMs, kernel density estimators, RBF-SVMs or Support Data Descriptors (SDD), may sometimes have bounded open space risk. However, any detector that uses linear classifiers, HAAR cascades, or Softmax-based classifiers, will almost always have an unbounded open set risk, and hence does not solve OSR.

"What about classifier with rejection?" you might ask. "Doesn't that solve the OSR problem?" The problem of rejecting some inputs in learning is more than 60 years old, (Chow 1957). Since its earliest days, the focus of rejecting has been the ambiguity between classes, not for addressing unknown inputs. As (Chow 1970) put it "The option to reject is introduced to safeguard against excessive misrecognition; it converts potential misrecognition into rejection." Chow's paper derives optimal thresholds for that multi-class recognition, assuming probabilities of classes are given over the whole feature space. The paper's thresholds were for optimizing the ambiguous regions between classes – it had the infinite regions of acceptance and infinite regions of rejection. Uncertainty is high near the decision boundary and most classifiers increase confidence with distance from the decision boundary. Thus, an unknown far from the boundary is not only incorrectly labeled but will be incorrectly classified with very high confidence. Even the majority of recent classifiers with reject-options (Fumera and Roli 2002; Rong and Metaxas 2006; Grandvalet et al. 2008; Bartlett and Wegkamp 2008; Wegkamp and Yuan 2011), all have the same issues and formally don't solve OSR problems because they can and generally have infinite positively labeled open space, hence infinite open space risk. The only classification with rejection options that do solve OSR are those that also satisfy the CAP criterion discussed above, e.g., a K-nearest neighbor with a threshold on an appropriate measure (Li and Wechsler 2005) or threshold on one or more one-class classifiers (Tax and Duin 2008; Liu et al. 2016a), in which cases there exists a threshold that will produce finite open space risk.

Novelty and anomaly detection algorithms (Bodesheim et al. 2015; Lazzaretti et al. 2016; Schultheiss et al. 2017; Bansal and Weld 2018) are solving different but related problems. Most such techniques apply a distributional model that is thresholded to detect the anomaly. They bound open space risk; they inspired our CAP model. However, their formulations do not balance open space risk with multi-class recognition risk as in Eq 1. They are only OSR for one-class problems. We can summarize the relationship informally as:

OSR $\approx$ Novelty-detection + multi-class recognition.

5 Open Set Deep Networks

The problem of open recognition is well formulated in either image or feature space, but people normally think of the world in image space. The above formal work on OSR was all done in "explainable" feature spaces, i.e., there existed a well-behaved mapping between input and the feature spaces. Due to the large complexity and the black box nature of deep networks, this mapping is not well-behaved. Moreover, their increased applications in the real world open them to a vast majority of unknown inputs, hence addressing OSR for deep networks is essential. We follow the OSR definition and just deal with the problem of OSR in testing, leaving for future work that this ignores the potential problem of unknowns showing up in the training data.

While rejecting unknown inputs by thresholding the network score is common(Matan et al. 1990; De Stefano, Sansone, and Vento 2000), thresholding softmax is problematic. Almost since its inception (Bridle 1989), softmax has been known to bias the probabilities towards a particular class even though the difference between the logit values of the winning class and other classes is minimal. This was highlighted by (Matan et al. 1990) who note that softmax would increase scores for a particular class even though they may have very limited activation on the logit level. In order to train the network to provide better logit values, they include an additional parameter α in the softmax loss by modifying the loss function as: $S_c = \log e^{l_c} / \left(e^{\alpha} + \sum_{c'=1}^{C} e^{l_{c'}}\right)$. During training, this forces a higher loss when the logit values, l, are smaller than α , and decrease the softmax scores when all logit values were smaller than α. This additional parameter can be interpreted as an additional node in the output layer that is not updated during back-propagation. Note, however, that like the other rejection approaches, this carves out a small region around the origin, and potentially between two classes as "none of the above," and still leaves infinite open space risk and hence does not solve OSR. In addition, there is also active research in network *uncertainty estimation* (Gal and Ghahramani 2016; Lakshminarayanan, Pritzel, and Blundell 2017; Mor and Wolf 2018). The authors of such claim thresholding their uncertainity can reject outliers. While more advanced than (Matan et al. 1990), these still suffer from infinite open space risk and hence do not solve OSR.

Another decades-old rejection approach used in neural networks is to add a "garbage" class (Linden and Kindermann 1989) or "background" class (Chang and Lippmann 1994). The "background" class is more effective than just thresholding uncertainty and hence is part of almost all modern multi-class detector such as (Liu et al. 2016b; Girshick 2015; Ren et al. 2015; Zhang et al. 2018). It must be noted that this background classifier just adds another class to reject instances not belonging to any of the known classes, but leaves infinite open space beyond each of the known classes. For different Bayesian Neural Networks (Ghosh, Delle Fave, and Yedidia 2016), like any Bayesian-based approach, makes a closed world assumption for all probability computations and is not suitable for OSR.

Background-class-based approaches capture some of the "known unknowns." However, background approaches do not

limit the space to have finite positively labeled open space, and hence they are formally not OSR. Background class modeling can work pretty well for self-consistent datasets like PASCAL (Everingham et al. 2010) and MS-COCO (Lin et al. 2014), where algorithms are often evaluated. However, the non-OSR property is a likely source of "negative" dataset bias (Tommasi et al. 2017) and limits application in the real world where the negative space has near infinite variety of inputs that should be rejected.

The OpenMax approach (Bendale and Boult 2016) is the first deep network approach to solve OSR formally. OpenMax uses EVT to define a per-class CAP model around the deep features of each class and then thresholds that probability to reject unknown inputs. The paper uses the penultimate layer, the logits, as a high-dimensional deep feature vector. However, the OpenMax approach could be used with any deep feature layer. Using the EVT model built from the positive instances training samples, it builds a per-class EVT probabilistic model of the input not belonging to that class, combining these in the OpenMax estimate of each class probability including the probability of it being unknown. Though this approach provides the first steps to formally address the open set issue for deep networks, it is an offline solution after the network had already been trained. Our more recent EVM model (Rudd et al. 2018), also provides a non-linear radial basis function approach based on EVT, but it provides a more flexible and more powerful representation model for OpenMax than the single class mean used in the original OpenMax paper. Neither of these models, however, are used in training the deep features.

One can see that a few networks with distance-based loss functions, such as center loss (Wen et al. 2016) and its variants, can be converted into an OSR network by simple thresholding on Euclidean distance from each class center as well as by applying an OpenMax layer.

More recent work (Dhamija, Günther, and Boult 2018) seeks to train deep networks that handle unknowns by combining softmax with a new loss function which forces known unknown samples, i.e., background classes, to have a small feature magnitude. While empirically the approach is considerably better than either using a background class or OpenMax, it formally has an unbounded open space risk.

The final OSR issue related to deep networks is related to the question of what one means by being far from training data. For traditional features, it was generally fine to think about it in either input space or feature space. However, for deep networks, adversarial examples (Szegedy et al. 2014) and fooling images (Nguyen, Yosinski, and Clune 2015) clearly break that parallelism. Both of these approaches produce images that are close in one space and far in the other. Moreover, while OpenMax was able to detect and mitigate most of the fooling images and simple adversarial examples, using an attack that goes after deep features rather than the end-to-end network we can manipulate just about any image so that it matches the features of a target and were easily able to defeat OpenMax (Rozsa, Günther, and Boult 2017). As of this writing, we are unaware of any defense against this attack. Thus, for deep networks, while we can develop OSR in feature space, it is not clear how to make them robust in image space, which is, of course, where the real world projects.

6 Conclusion

While machine learning and deep networks are providing great advances, and many application areas are awash with data, no amount of training will prepare the system for all unknown inputs. Rejecting uncertain inputs is not enough — uncertain is not unknown, and unknowns are often labeled with confidence. Novelty, anomaly or outlier detection alone is not sufficient because the proper handling of unknowns involves balancing the risk of the unknown with the risk from recognition errors. With almost all deep networks, unknowns map into the same space as knowns and are not easily rejected as outliers.

While significant progress has been made, as the astute reader who was checking references while reading may have noted, more than a dozen of the references have titles like "Towards..."; this is an emerging area with more unknowns than knowns. Open set problems are often challenging because they must balance maintaining accuracy on the core problem with handling the unknown unknowns. However, dealing with the unknown is essential, and we need systems explicitly designed to handle the unknown. **Do not fear the unknown — join us in taming it.**

> *"For any scientist, the real challenge is not to stay within the secure garden of the known but to venture out into the wilds of the unknown." – (Du Sautoy 2017)*

References

Bansal, G., and Weld, D. S. 2018. A coverage-based utility model for identifying unknown unknowns. In *Proc. of AAAI*.

Bao, J.; Chen, D.; Wen, F.; Li, H.; and Hua, G. 2018. Towards open-set identity preserving face synthesis. In *IEEE CVPR*.

Bapst, A. B.; Tran, J.; Koch, M. W.; Moya, M. M.; and Swahn, R. 2017. Open set recognition of aircraft in aerial imagery using synthetic template models. In *Automatic Target Recognition XXVII*, volume 10202, 1020206.

Bartlett, P., and Wegkamp, M. 2008. Classification with a reject option using a hinge loss. *In JMLR* 9:1823–1840.

Battaglino, D.; Lepauloux, L.; and Evans, N. 2016. The open-set problem in acoustic scene classification. In *2016 IEEE Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 1–5. IEEE.

Bendale, A., and Boult, T. 2015. Towards open world recognition. In *IEEE CVPR*, 1893–1902.

Bendale, A., and Boult, T. E. 2016. Towards open set deep networks. In *IEEE CVPR*, 1563–1572.

Bodesheim, P.; Freytag, A.; Rodner, E.; and Denzler, J. 2015. Local novelty detection in multi-class recognition problems. In *IEEE WACV*, 813–820.

Bridle, J. 1989. Probablistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. *Neuro-computing: Algorithms, Architectures*.

Busto, P. P., and Gall, J. 2017. Open Set Domain Adaptation. In *ICCV*, 754–763.

Caragea, D.; Silvescu, A.; and Honavar, V. 2000. Incremental and distributed learning with support vector machines. In *AAAI*, 1067.

Cardoso, D. O.; França, F.; and Gama, J. 2015. A bounded neural network for open set recognition. In *Neural Networks (IJCNN), 2015 Int. Joint Conf. on*, 1–7. IEEE.

Chang, E. I., and Lippmann, R. P. 1994. Figure of merit training for detection and spotting. In *NIPS*, 1019–1026.

Chao, W.-L.; Changpinyo, S.; Gong, B.; and Sha, F. 2016. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In *ECCV*, 52–68. Springer.

Chen, Z., and Liu, B. 2016. Lifelong machine learning. *Synthesis Lect. on Art. Int. and Machine Learning* 10(3):1–145.

Chiachia, G.; Falcao, A. X.; Pinto, N.; Rocha, A.; and Cox, D. 2014. Learning person-specific representations from faces in the wild. *IEEE TIFS* 9(12):2089–2099.

Chow, C.-K. 1957. An optimum character recognition system using decision functions. *IRE Trans. on Ele. Comp.* 4:247–254.

Chow, C. 1970. On optimum recognition error and reject tradeoff. *IEEE Trans. Info. Theory* 16(1):41–46.

Costa, F. d. O.; Silva, E.; Eckmann, M.; Scheirer, W. J.; and Rocha, A. 2014. Open set source camera attribution and device linking. *Pat. Rec. Letters* 39:92–101.

Cruz, S.; Coleman, C.; Rudd, E. M.; and Boult, T. E. 2017. Open set intrusion recognition for fine-grained attack categorization. In *IEEE Tech. for Homeland Security*.

De Stefano, C.; Sansone, C.; and Vento, M. 2000. To reject or not to reject: that is the question-an answer in case of neural classifiers. *IEEE TSMC, Part C* 30(1):84–94.

Dhamija, A.; Günther, M.; and Boult, T. E. 2018. Reducing network agnostophobia. In *NIPS*.

Dietterich, T. G. 2017. Steps toward robust artificial intelligence. *AI Magazine* 38(3):3–24.

Doan, T., and Kalita, J. 2017. Overcoming the challenge for text classification in the open world. In *Computing and Communication Workshop and Conf.* IEEE.

Du Sautoy, M. 2017. *The Great Unknown: Seven Journeys to the Frontiers of Science*. Penguin.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *Int. Journal of Computer Vision* 88(2):303–338.

Ferreira, A., and Giraldi, G. 2017. Convolutional Neural Network approaches to granite tiles classification. *Expert Systems with Applications* 84:1–11.

Fu, Y.; Xiang, T.; Jiang, Y.-G.; Xue, X.; Sigal, L.; and Gong, S. 2018. Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Processing Magazine* 35(1):112–125.

Fumera, G., and Roli, F. 2002. Support vector machines with embedded reject option. In *Pat. Rec. with Support Vector Machines*. Springer. 68–82.

Gal, Y., and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Int. Conf. on Machine Learning*, 1050–1059.

Ghosh, S.; Delle Fave, F. M.; and Yedidia, J. S. 2016. Assumed density filtering methods for learning bayesian neural networks. In *AAAI*, 1589–1595.

Girshick, R. 2015. Fast R-CNN. In *IEEE ICCV*, 1440–1448.

Gopalan, R.; Li, R.; Patel, V. M.; and Chellappa, R. 2015. Domain adaptation for visual recognition. *Foundations and Trends®in Computer Graphics and Vision* 8(4):285–378.

Grandvalet, Y.; Rakotomamonjy, A.; Keshet, J.; and Canu, S. 2008. Support vector machines with a reject option. In *NIPS*, 537–544. Curran Associates, Inc.

Grave, E.; Cisse, M. M.; and Joulin, A. 2017. Unbounded cache model for online language modeling with open vocabulary. In *NIPS*, 6042–6052.

Günther, M.; Hu, P.; Herrmann, C.; Chan, C. H.; Jiang, M.; Yang, S.; Dhamija, A. R.; Ramanan, D.; Beyerer, J.; Kittler, J.; et al. 2017. Unconstrained face detection and open-set face recognition challenge. In *IEEE Int. Joint. Conf. Biometrics*, 697–706.

Günther, M.; Cruz, S.; Rudd, E. M.; and Boult, T. E. 2017. Toward open-set face recognition. In *IEEE CVPR Workshops*.

Henrydoss, J.; Cruz, S.; Rudd, E. M.; and Boult, T. E. 2017. Incremental Open Set Intrusion Recognition Using Extreme Value Machine. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE Int. Conf. on*, 1089–1093. IEEE.

Hewitt, C., and Jong, P. D. 1983. Analyzing the roles of descriptions and actions in open system. In *Proceedings of AAAI-83*, 162–167. See also AI Lab Memo727, http://www.dtic.mil/dtic/tr/fulltext/u2/a133614.pdf.

Hodge, V. J., and Austin, J. 2004. A survey of outlier detection methodologies. *Artificial Intelligence Review* 22(2):85–126.

Jain, L. P.; Scheirer, W. J.; and Boult, T. E. 2014. Multi-class open set recognition using probability of inclusion. In *ECCV*.

Juefei-Xu, F., and Savvides, M. 2016. Multi-class Fukunaga Koontz discriminant analysis for enhanced face recognition. *Pat. Rec.* 52:186–205.

Júnior, P. R. M.; Boult, T. E.; Wainer, J.; and Rocha, A. 2016. Specialized support vector machines for open-set recognition. *arXiv:1606.03802*.

Krstulović, S. 2018. Audio event recognition in the smart home. In *Computational Analysis of Sound Scenes and Events*. Springer. 335–371.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS*, 6405–6416.

Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2014. Attribute-based classification for zero-shot visual object categorization. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 36(3):453–465.

Lazzaretti, A. E.; Tax, D. M. J.; Neto, H. V.; and Ferreira, V. H. 2016. Novelty detection and multi-class classification in power distribution voltage waveforms. *Expert Systems with Applications* 45:322–330.

Li, F., and Wechsler, H. 2005. Open set face recognition using transduction. *IEEE transactions on pattern analysis and machine intelligence* 27(11):1686–1697.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *ECCV*, 740–755. Springer.

Linden, A., and Kindermann, J. 1989. Inversion of multilayer nets. In *Proc. Int. Joint Conf. Neural Networks*, volume 2, 425–430.

Liu, J.; Miao, Q.; Sun, Y.; Song, J.; and Quan, Y. 2016a. Modular ensembles for one-class classification based on density analysis. *Neurocomputing* 171:262–276.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; and Berg, A. C. 2016b. SSD: Single shot multibox detector. In *ECCV*, 21–37. Springer.

Markou, M., and Singh, S. 2003. Novelty detection: a review-part 1: statistical approaches. *Signal Processing* 83(12):2481–2497.

Matan, O.; Kiang, R.; Stenard, C.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L.; and Le Cun, Y. 1990. Handwritten character recognition using neural network architectures. In *4th USPS Adv. Technology Conf.*, 1003–1011.

Mensink, T.; Verbeek, J.; Perronnin, F.; and Csurka, G. 2012. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In *ECCV*. Springer.

Mor, N., and Wolf, L. 2018. Confidence prediction for lexicon-free OCR. In *IEEE WACV*.

Mu, X.; Ting, K. M.; and Zhou, Z.-H. 2017. Classification under streaming emerging new classes: A solution using completely-random trees. *IEEE Trans. on Knowledge and Data Engineering* 29(8):1605–1618.

Navarro, L. C.; Navarro, A. K.; Rocha, A.; and Dahab, R. 2018. Connecting the dots: Toward accountable machine-learning printer attribution methods. *Journal of Visual Communication and Image Representation* 53:257–272.

Neal, L.; Olson, M.; Fern, X.; Wong, W.-K.; and Li, F. 2018. Open set learning with counterfactual images. In *ECCV*, 613–628.

Neira, M. A. C.; Júnior, P. R. M.; Rocha, A.; and Torres, R. D. S. 2018. Data-Fusion Techniques for Open-Set Recognition Problems. *IEEE Access* 6:21242–21265.

Nguyen, A.; Yosinski, J.; and Clune, J. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE CVPR*.

Perera, P., and Patel, V. M. 2017. Extreme value analysis for mobile active user authentication. In *2017 12th IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG 2017)*, 346–353. IEEE.

Pinto, A.; Schwartz, W. R.; Pedrini, H.; and de Rezende Rocha, A. 2015. Using visual rhythms for detecting video-based facial spoof attacks. *IEEE Trans. on Information Forensics and Security* 10(5):1025–1038.

Poitevin, P.; Pelletier, M.; and Lamontagne, P. 2017. Challenges in detecting UAS with radar. In *Security Technology (ICCST), 2017 Int. Carnahan Conf. on*. IEEE.

Prakhya, S.; Venkataram, V.; and Kalita, J. 2017. Open-Set Deep Learning for Text Classification. *Machine Learning in Computer Vision and NLP*.

Rattani, A.; Scheirer, W. J.; and Ross, A. 2015. Open set fingerprint spoof detection across novel fabrication materials. *IEEE Trans. on Information Forensics and Security* 10(11):2447–2460.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proc. CVPR*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 91–99.

Rocha, A.; Scheirer, W. J.; Forstall, C. W.; Cavalcante, T.; Theophilo, A.; Shen, B.; Carvalho, A. R.; and Stamatatos, E. 2017. Authorship attribution for social media forensics. *IEEE Trans. on Information Forensics and Security* 12(1):5–33.

Rong, Z., and Metaxas, D. 2006. RO-SVM: Support vector machine with reject option for image categorization. In *BMVC*.

Roos, J. D., and Shaw, A. K. 2017. Probabilistic SVM for open set automatic target recognition on high range resolution radar data. In *Automatic Target Recognition XXVII*, volume 10202, 102020B.

Rozsa, A.; Günther, M.; and Boult, T. E. 2017. Adversarial robustness: Softmax versus openmax. In *BMVC17*. arXiv:1708.01697.

Rudd, E.; Rozsa, A.; Gunther, M.; and Boult, T. 2017. A survey of stealth malware: Attacks, mitigation measures, and steps toward autonomous open world solutions. *IEEE Communications Surveys & Tutorials* 19(2):1145–1172.

Rudd, E. M.; Jain, L. P.; Scheirer, W. J.; and Boult, T. E. 2018. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence* 40(3):762–768.

Scheirer, W. J.; Rocha, A.; Sapkota, A.; and Boult, T. E. 2013. Towards open set recognition. *IEEE T-PAMI* 36.

Scheirer, W.; Jain, L.; and Boult, T. 2014. Probability models for open set recognition. *IEEE T-PAMI* 36:2317–2324.

Scherreik, M., and Rigling, B. 2016a. Multi-class open set recognition for SAR imagery. In *Automatic Target Recognition XXVI*, volume 9844, 98440M. Int. Society for Optics and Photonics.

Scherreik, M. D., and Rigling, B. D. 2016b. Open set recognition for automatic target classification with rejection. *IEEE Trans. on Aerospace and Electronic Systems* 52(2):632–642.

Schraml, R.; Debiasi, L.; Kauba, C.; and Uhl, A. 2017. On the feasibility of classification-based product package authentication. In *IEEE Information Forensics and Security (WIFS)*. IEEE.

Schultheiss, A.; Käding, C.; Freytag, A.; and Denzler, J. 2017. Finding the Unknown: Novelty Detection with Extreme Value Signatures of Deep Neural Activations. In *German Conf. on Pattern Recognition*, 226–238. Springer.

Szegedy, C. J.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; and Fergus, R. 2014. Intriguing properties of neural networks. In *Int. Conf. on Learning Representation (ICLR)*.

Sünderhauf, N.; Dayoub, F.; McMahon, S.; Talbot, B.; Schulz, R.; Corke, P.; Wyeth, G.; Upcroft, B.; and Milford, M. 2016. Place categorization and semantic mapping on a mobile robot. In *Robotics and Automation (ICRA), IEEE Int. Conf. on*, 5729–5736. IEEE.

Tax, D. M., and Duin, R. P. 2008. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters* 29(10):1565–1570.

Tommasi, T.; Patricia, N.; Caputo, B.; and Tuytelaars, T. 2017. A deeper look at dataset bias. In *Domain Adaptation in Computer Vision Applications*. Springer. 37–55.

Wegkamp, M., and Yuan, M. 2011. Support vector machines with a reject option. *Bernoulli* 17(5):1368–85.

Wen, Y.; Zhang, K.; Li, Z.; and Qiao, Y. 2016. A discriminative feature learning approach for deep face recognition. In *ECCV*, 499–515.

Xian, Y.; Lampert, C. H.; Schiele, B.; and Akata, Z. 2018. Zero-shot learning-A comprehensive evaluation of the good, the bad and the ugly. *IEEE TPAMI*.

Xian, Y.; Schiele, B.; and Akata, Z. 2017. Zero-shot learning-the good, the bad and the ugly. In *IEEE CVPR*, 3077–3086.

Zamora, E., and Yu, W. 2016. Novel autonomous navigation algorithms in dynamic and unknown environments. *Cybernetics and Systems* 47(7):523–543.

Zhang, H., and Patel, V. M. 2017. Sparse representation-based open set recognition. *IEEE transactions on pattern analysis and machine intelligence* 39(8):1690–1696.

Zhang, S.; Benenson, R.; Omran, M.; Hosang, J.; and Schiele, B. 2018. Towards reaching human performance in pedestrian detection. *IEEE TPAMI* 40(4):973–986.